Transforming Adolescent Healthcare in Rwanda: Sustainable and Scalable Features in Digital Health

Tejaswi Samrat Dasari*, Pratima Satish†, Laetitia Kayitesi‡, Mary Mbuvi‡, Russell Dias‡,
Fabiola Ishimwe Ngamije‡, Joseph Kwesiga‡, Robert On§, Diane Myung-kyung Woodbridge*
*University of San Francisco, † University of California, Berkeley
‡YLabs, §The Agency Fund
{tdasari, dwoodbridge}@usfca.edu, pratimasatish@berkeley.edu,
{laetitia.kayitesi, mary.mbuvi, fabiola.ngamije, joe.kwesiga}@ylabsglobal.org, robert@agency.fund

***Abstract*—CyberRwanda is a digital health intervention designed to increase knowledge of family planning and reproductive health (FP/RH) and access to youth-friendly services in Rwanda. Previous studies showed high acceptability and feasibility through user surveys, success in combating social stigma, and an effective educational impact on adolescents. In order to further enhance user experience and overcome resource limitations, the authors designed and developed new features, including an integrated dashboard and chatbot, leveraging data engineering and natural language processing. The experiments showed that new features substantially reduce the time required to access information for user groups, potentially leading to an improved user experience.**

***Index Terms*—Big Data applications, Data engineering, Natural language processing, Public healthcare, Smart healthcare**

I. Introduction

The Rwandan government offers comprehensive sexuality education in secondary schools to improve knowledge of puberty, reproduction, pregnancy, gender-based violence (GBV), sexual risk behavior, and HIV/sexually transmitted infections (STIs) [1]. However, several studies have shown that sexual education in Rwanda is challenged primarily due to the presence of stigma regarding premarital sex. These challenges have far-reaching implications for the well-being of individuals, as well as social and economic disparities, such as lower levels of education and limited employment opportunities for women [2] [3]. Notably, Rwanda exhibits a higher HIV prevalence of 3.0% among adults, surpassing the global average of 0.7%, with females being particularly affected [4] [5] [6]. Although the rate of adolescent pregnancy in Rwanda decreased from 7% in 2015 to 5% in 2020, there are still significant gaps in adolescent FP/RH services. According to the 2021 Demographic and Health Survey, approximately half of all unmarried girls aged between 15 and 19 lack access to modern contraception despite expressing a desire for it [7].

The CyberRwanda platform was developed by Youth Development Labs (YLabs) with the participation of over a thousand Rwandan youth, parents, teachers, health care providers, and community leaders using a human-centered design approach. CyberRwanda aims to improve access to health products and information on FP/RH and employment for Rwandan adolescents using storytelling, interactive FAQs, and an online shop for confidential ordering of health products from youth-friendly pharmacies. The results from a pilot study of CyberRwanda showed a high level of satisfaction among participants, who appreciated the platform's ability to create a safe and confidential environment for seeking information and accessing essential products [8] [9].

CyberRwanda includes three main sections: STORIES (storytelling content on FP/RH and employment); LEARN (over 300 frequently asked questions (FAQs) and an advice column); SHOP (online store for youth to discreetly order health products). Through the LEARN section, youth are able to submit questions which are then answered by human experts. However, there were noticeable fluctuations in the number of questions asked to the expert teams, as depicted in Figure 1. These fluctuations were influenced by school openings and launch events, which impacted the overall wait time for questions being answered (Table II).

The existing CyberRwanda system architecture comprises heterogeneous databases responsible for collecting web analytics data from Google Analytics and tracking a user's order history in a pharmacy backend (Figure 2). To analyze platform usage and effectiveness, experts at YLabs were required to access each database and its corresponding frontend separately. This led to a fragmented approach that hindered a comprehensive understanding of the intervention outcomes. Additionally, the human experts had to log in to two different systems, adding complexity and inconvenience to the process.

In order to overcome the current limitations on the platform usage analysis, this study focuses on the implementation of a comprehensive dashboard that integrates data from diverse sources, providing the experts with a unified platform to visualize and analyze trends. Additionally, to improve user experience, a chatbot was developed to facilitate higher engagement by leveraging existing textual data and natural language processing (NLP) techniques.

II. System Overview

In this work, we have introduced two new features: an integrated dashboard and a chatbot to enhance the experience of the human experts at YLabs and the Rwandan youth, respectively. The integrated dashboard serves the purpose of consolidating data from multiple sources and presenting application usage trends as well as pharmacy orders. Additionally, we have created a question-answering chatbot that employs

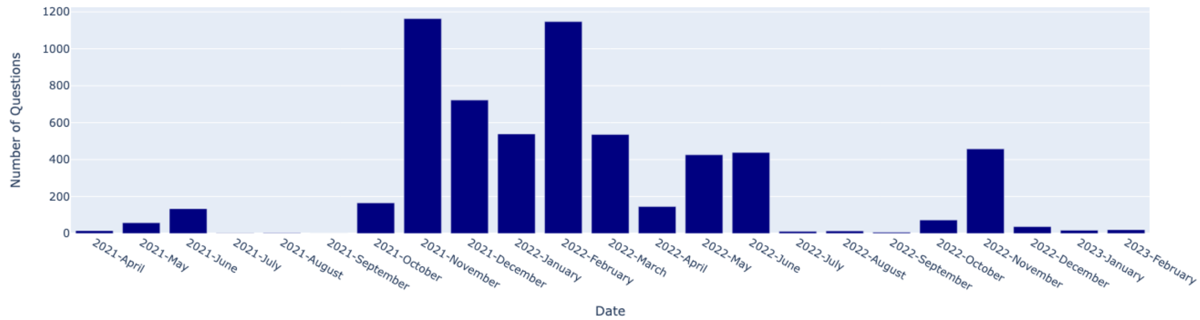


Fig. 1: Number of Questions Posted. Note that due to the pandemic and its lockdown measures, schools in Rwanda experienced intermittent closures between March and October 2021. However, schools were able to reopen on October 11th, 2021. The initial stages of CyberRwanda included a soft launch at the end of May, 2021. The official launch events took place in mid-October, 2021, and subsequent launches occurred in January, May, and October, 2022. Schools typically have exams and holiday breaks in March, July and December. Additionally, schools observe closures in April, and mid-July to September.

TABLE I: Data Sources and Visualisation Dashboard

Original Data Sources	**Visualisation**
Web Analytics	District Interactions
	Chapter Interactions
	User Engagement Time
	Device Category
	Web Application Distribution
Pharmacy Backend	Pharmacy Order Status
	Product Distribution
	Picked Products Distribution

NLP to respond to user questions. This helps minimize users' wait time and enables more efficient use of human experts' time, as the chatbot can provide answers if they can be inferred from the existing data.

A. *Integrated Dashboard*

To overcome the limitations of the existing infrastructure mentioned in Section I, we designed and developed an integrated dashboard that incorporates a streamlined backend pipeline (Figure 2). This pipeline facilitates the integration of data from both Google Analytics and pharmacy backend MongoDB databases, enabling a unified view.

Furthermore, the extended framework offers a user-friendly, customizable dashboard with eight different graphs, each providing valuable insights into the CyberRwanda platform's performance (Table I). Human experts can easily select a desired date range to visualize data using interactive graphs, thereby empowering them to explore trends and patterns effectively. To facilitate data analysis and further investigation, the dashboard also incorporates a download feature to download data in CSV format and the visualizations as interactive image files.

B. *Question-answering Chatbot*

The objective of this application is to find the best response from existing FAQs in Kinyarwanda/Swahili that matches the user's query. The FAQ database contains 300 questions and their corresponding answers in English and Kinyarwanda, focusing on the health and well-being of adolescents. Each entry in the database includes a slug, which represents the word split of a question. For instance, 'how-can-i-develop-effective-study-habits' is the slug for "How can I develop effective study habits?".

We applied NLP techniques to find the top k FAQs that are most relevant to the end user's query and reduce the response wait time. Figure 3 shows the data pipeline of the chatbot.

1) Language Detection and Translation: The present task addresses the challenge of processing user queries in diverse African languages to translate into languages where the majority of the language models are built, which is English. The downstream embedding model employed in this study is specifically designed for English language input. Therefore, the translation of user queries into English is imperative. To address this issue, the Google Cloud Translate API [10] is employed as it can detect the language of the user query from any of the African languages and subsequently translate it into the desired destination language. This feature provides the potential to empower developers to create similar applications for other countries, utilizing general information and eliminating language barriers.

2) Encoding FAQs and User Query: The encoding task involves representing the translated text into a vector in a 768-dimensional latent space to capture rich contextual information and preserve the word order within a given text.

This is accomplished using a fine-tuned version of Masked and Permuted Pre-training Model for Language Understanding (MPNet) [11]. The model is specifically designed to encode sentences and transform them into fixed-length numerical representations called embeddings and has been trained on a vast and diverse dataset consisting of over 1 billion training pairs. MPNet combines the benefits of BERT [12] and XLNet [13] while overcoming their limitations, utilizing permuted

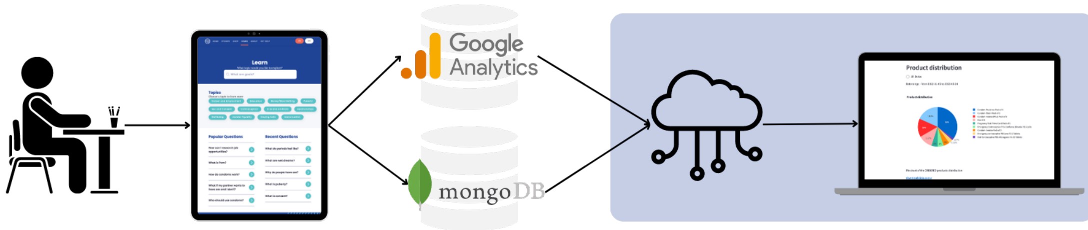


Fig. 2: System Overview - The extended framework (highlighted) includes a backend pipeline for integrating data from heterogeneous databases including Google Analytics and Pharmacy backend database to create a frontend dashboard.

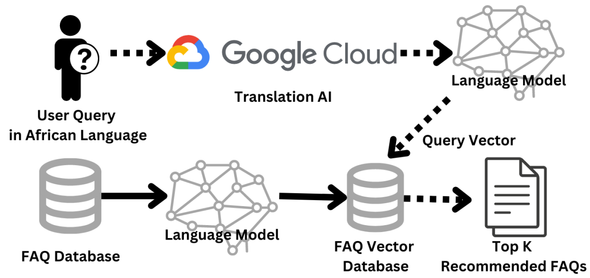


Fig. 3: Question-answering Chatbot Design

language modeling to capture token dependencies and incorporating full position information to reduce position discrepancy.

Each question in the FAQs database is encoded using this sentence transformer model, resulting in a corresponding vector representation. These vectors are then indexed for efficient search and retrieval.

3) Answer Recommendation: After encoding a user query into a vector representation, we perform a search operation on the FAQs database, seeking similar questions and their answers. The cosine similarity metric is employed to quantify the similarity between the query vector and the vectors representing the question strings in the FAQs database. As cosine similarity ranges from -1 to +1, it provides a more intuitive understanding and comparison in matching the potential FAQs compared to Euclidean distance.

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\sum_{i=1}^{n} \mathbf{x}_i \mathbf{y}_i}{\sqrt{\sum_{i=1}^{n} (\mathbf{x}_i)^2} \sqrt{\sum_{i=1}^{n} (\mathbf{y}_i)^2}} \tag{1}$$

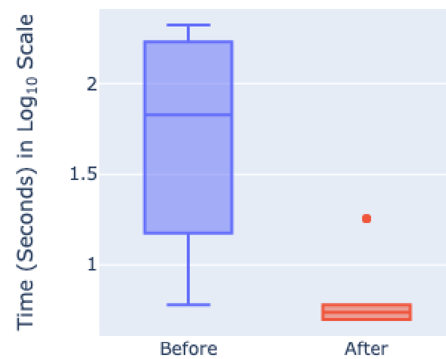
Further, we retrieve the top k similar FAQs based on their cosine similarity scores. Each FAQ is accompanied by its corresponding English version as well as its African language counterpart, with an option to provide the end users with either of the responses or both.

III. Result

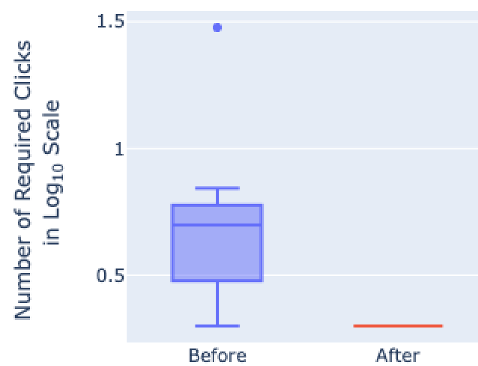
To evaluate the efficiency of the new features, we compared the required amount of time and user interactions.

A. Integrated Dashboard

The experimental results focusing on the efficiency of the new integrated dashboard in terms of time and the number



(a) Time required for accessing data



(b) Number of clicks required for accessing data

Fig. 4: Comparison between before and after implementing the integrated dashboard

of clicks required to access data are presented in Figure 4. To evaluate the statistical significance of the differences, a Mann-Whitney U test was performed. The test yielded a p-value of 2.70×10^{-6} and 3.68×10^{-6}, respectively, for time and the number of clicks. This indicates a significant improvement in both time efficiency and user interaction with the integrated dashboard compared to the previous system.

B. Question-answering Chatbot

Table II presents the response times for answering user questions. This shows that the number of asked questions (Figure 1) is significantly larger than the number of questions

TABLE II: Response time (hours) for user questions between April 2021 and February 2023

Year	Month	Answered Questions	Response Time (Hours)		
			25%	50%	75%
2021	10	3	8724.5	8785.0	9038.0
	11	31	4160.0	4690.0	7939.5
2022	11	4	2.5	6.0	10.5
	12	1	392.0	392.0	392.0
2023	1	1	113.0	113.0	113.0
	2	1	24.0	24.0	24.0

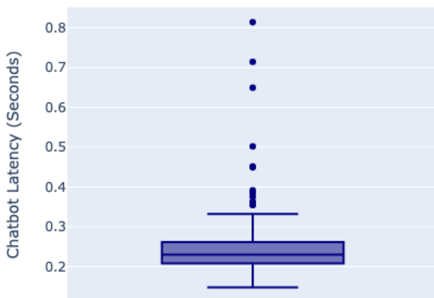


Fig. 5: Latency of chatbot system for randomly selected 106 questions.

for which responses are provided. The response time varies due to several factors, such as the types and quality of questions, as well as the availability of human resources. Notably, questions requiring specialized medical expertise may take longer to be answered, given the specific knowledge and skills needed for accurate responses.

To assess the time efficiency of the newly developed chatbot, a sample of 106 questions was randomly selected. These questions were inputted into the chatbot, and we received responses within one second for all queries, where the median response time was 0.23 seconds (Figure 5). This shows that the chatbot can help reduce the burden on human experts and minimize wait times for users, leading to the enhancement of the overall user experience.

IV. Conclusion

In this study, the authors designed and productized features aimed at enhancing the user experience of CyberRwanda, a digital health platform for adolescents in Rwanda. By incorporating an integrated dashboard and chatbot functionality, the authors successfully reduced the time required for accessing the information on both the human experts' and users' ends, thereby improving overall efficiency and access to information. The new dashboard enables experts to see detailed system usage and various business metrics for decision-making processes, by integrating data from multiple platforms and reducing the number of steps and time to produce reports. On the product user side, the chatbot feature effectively addresses the need for timely sexual education information, minimizing wait times for adolescent youth and contributing to the platform's long-term sustainability. The experiment outcomes demonstrated improved time efficiency, directly correlating with a positive user experience.

In future experiments, the researchers will assess the extended impact of the features, examining their effects on health outcomes and cost-effectiveness. In addition, the authors plan to expand their investigation to incorporate additional data sources, such as medical journals and databases, to enhance the chatbot's ability to generate accurate and comprehensive answers. Furthermore, the future features should be able to consider the unique characteristics of each region, including cultural nuances and available resources, to ensure that the provided information remains appropriate and relevant to the specific needs of each region. By continuing to refine and expand the capabilities of CyberRwanda, the authors seek to contribute to the ongoing advancement of digital health platforms and improve the sexual education and well-being of adolescents in Rwanda and beyond.

References

[1] Ministry of Education, Republic of Rwanda, "Competence-based curriculum, curriculum framework - pre-primary to upper secondary," 2015.

[2] A. Nkurunziza, N. Van Endert, J. Bagirisano, J. B. Hitayezu, S. Dewaele, O. Tengera, and G. Jans, "Breaking barriers in the prevention of adolescent pregnancies for in-school children in kirehe district (rwanda): a mixed-method study for the development of a peer education program on sexual and reproductive health," *Reproductive Health*, vol. 17, no. 1, pp. 1–8, 2020.

[3] H. Schwandt, A. Boulware, J. Corey, A. Herrera, E. Hudler, C. Imbabazi, I. King, J. Linus, I. Manzi, M. Merritt *et al.*, ""she is courageous because she does not care what people think about her…": attitudes toward adolescent contraception use among rwandan family planning providers and adult female modern contraceptive users," *Reproductive Health*, vol. 19, no. 1, p. 204, 2022.

[4] S. Nsanzimana, G. N. Rwibasira, S. S. Malamba, G. Musengimana, E. Kayirangwa, S. Jonnalagadda, E. F. Rezende, J. W. Eaton, V. Mugisha, E. Remera *et al.*, "Hiv incidence and prevalence among adults aged 15-64 years in rwanda: Results from the rwanda population-based hiv impact assessment (rphia) and district-level modeling, 2019," *International Journal of Infectious Diseases*, vol. 116, pp. 245–254, 2022.

[5] S. Nsanzimana, V. Mugisha, A. Guhathakurta, and D. Haroz, "Survey results put rwanda on track to achieve hiv epidemic control," Jan 2019. [Online]. Available: https://www.cdc.gov/media/releases/2019/p0925-survey-rwanda-hiv-control.html

[6] World Health Organization, "The global health observatory - hiv," 2021. [Online]. Available: https://www.who.int/data/gho/data/themes/hiv-aids

[7] The Demographic and Health Surveys (DHS) Program, "Rwanda demographic and health survey 2014-15," 2016.

[8] YLabs, CyberRwanda , "Using digital education to improve the health and livelihoods of urban adolescents," Aug 2022. [Online]. Available: https://www.ylabsglobal.org/work/all/cyberrwanda

[9] R. Hémono, L. Packel, E. Gatare, L. Baringer, N. Ippoliti, S. I. McCoy, and R. Hope, "Digital self-care for improved access to family planning and reproductive health services among adolescents in rwanda: preliminary findings from a pilot study of cyberrwanda," *Sexual and Reproductive Health Matters*, vol. 29, no. 3, p. 2110671, 2022.

[10] Google Cloud, "Translation AI," 2023. [Online]. Available: https://cloud.google.com/translate

[11] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, "Mpnet: Masked and permuted pre-training for language understanding," *arXiv preprint arXiv:2004.09297*, 2020.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[13] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Advances in neural information processing systems*, vol. 32, 2019.